Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 2 – No.2, June 2015 – www.caeaccess.org

Rough Set Approach for Generation of Classification Rules for Hepatitis

Sujogya Mishra
Research scholar,
Utkal University
Bhubaneswar-751004, India

Shakti Prasad Mohanty
Department of Mathematics,
College of Engineering and Technology
Bhubaneswar-751003, India

Sateesh Kumar Pradhan
Department of Computer Science,
Utkal University
Bhubaneswar-751003, India

Radhanath Hota
Department of Computer Science,
College of Basic Science and Hum
OUAT, Bhubaneswar-751003, India

ABSTRACT
In the current age research in the field of medical science has been increased to a significant height but there are several new virus which cannot be detect by the usual medical test , for example some common disease like malaria ,dengue, hepatitis , jaundice needs of very meticulous medical analysis because all the above said dieses has very common symptoms which needs of strong analysis to determine the exact dieses. Maximum number of medical test which are conducted to determine the dieses mostly based upon doctor's guess which are not only expensive and but also give inaccurate pathological result. In this paper we emphasized more on symptom rather than pathological test .From the large domain we consider the disease hepatitis for our purpose .Every year millions of people died from hepatitis due to improper diagnosis .We develop an algorithm using rough set concept to counter hepatitis. We classified the entire paper in to three basic section 1st section about literature review the 2nd and 3rd section deals with the Experiment, Findings, and Statistical Validation.

Keywords
Rough Set Theory, Medical related data, Granular computing, Data mining.

1. INTRODUCTION
The increasing growth of medical science and different types of un-common dieses always put challenge on doctors and scientist to counter. From the large medical database which creates a need and an opportunity to ex-tract inference from data bases[1] .Data base regarding human biology have gathered large quantities of information about the patients and their physical conditions, analysis of these data provided with new medical information often resulted incomplete information [2] .Often data analysis based upon assumptions of knowledge aboutdependencies, probability theory and large number of experiments, unable to de-rive correct conclusions neither from incomplete information nor manage the data consistency. The general intelligence techniques used in medical data analysis are Neural network[3] Bayesian classifier [4] Genetic algorithms[5] Decision trees [6] Fuzzy set [7] . Rough set theory and it's basic concept was invented by Polish logician, Professor Z. Pawlak in early eighties[8] .Rough set theory is based upon conventional set theory very useful for extracting knowledge from un-certain and incomplete data based information, it assumes that we first have necessary information or knowledge of all the objects in the universe with which the objects can be divided into different groups. If we have exactly the same information of two objects then we say that they are indiscernible (similar), i.e., we cannot distinguish them with known knowledge. It can be used to find dependencies among the data set to evaluate the importance of attributes to find the pattern of data. Learn general decision-making principle , decrease all redundant objects and attributes and seek the minimum subset of attributes so as to attain satisfying categories. More-over, Rough set reduction algorithms is very useful in analyzing and approximate the decision classes using simplified patterns [9].Rough set theory become very popular among scientists around the world and the it is now one of the most growing intelligent data analysis tool. Unlike other research methods such as fuzzy set theory, Dempster–Shafer theory or statistical methods, rough set analysis requires no external parameters and uses only the information presented in the given data [10].This paper discusses how rough set theory can be used to analyze medical data, and for extracting information from a set of observed samples of the hepatitis data. Rough set reduction technique is applied to find all reducts of the data which contains the minimum subset as attributes that are associated with a class for further classification. We classified the entire paper in three section as follows 1st section contains the literature review like definition and of rough sets and elementary concept of rough set theory, correlation and some statistical validation techniques the 2ndsection consists of data analysis on medical data which we collected from DR P K Mishra M D and 3rd section contains the algorithm for rule generation for classification of jaundice and statistical validation of the rule that is generated in the form of algorithm the 4th section contains the conclusion part and the future work.

2. PRILIMINARIES
2.1 Rough set
Rough set theory as introduced by Z. Pawlak[8] is an

extension of conventional set theory that support approximations in decision making.

2.1.2 Approximation Space: An Approximation space is a pair (U , R) where U is a non empty finite set called the universe R is an equivalence relation defined on U.

2.1.3 Information System: An information system is a pair S = (U , A), where U is thenon-empty finite set called the universe, A is the non-empty finite set of attributes

2.1.4 Decision Table: A decision table is a special case of information systems S= (U , A= C U {d}), where d is not in C. Attributes in C are called conditional attributes and d is a designated attribute called the decision attribute

.2.1.5 Approximations of Sets: Let S = (U, R) be an approximation space and X be a subset of U.The lower approximation of X by R in S is defined as $\underline{RX}$= { e ε U | [e] ε X} and The upper approximation of X by R in S is defined as $\overline{RX}=\{e \in U/[e] \cap X \neq \phi\}$where [e] denotes the equivalence class containing e. A subset X of U is said to be R-definable in S if and only if $\overline{RX}=\underline{RX}$ .A set X is rough in S if its boundary set is nonempty.

2.2 Dependency of Attributes

Let C and D be subsets of A. We say that D depends on C in a degree k (0 ≦k ≦1) denoted by C →k D if K=y(C,D)= $\left|\frac{POS_C(D)}{|u|}\right|$ where $POS_C(D) = U \underline{C}(X)$, is called positive region of the partition U/D with respect to C where $x \in u/d$, which is all elements of U that can be uniquely classified to the block of partition U/D. If k = 1 we say that D depends totally on C. If k < 1 we say that D depends partially (in a degree k) on C.

2.3 Dispensable and Indispensable Attributes

*Let S = (U, A = C ∪ D) be a decision table.*Let c be an attribute in C. Attribute c is dispensable in S if $POS_C(D) = POS_{(C-\{c\})}(D)$otherwise, c is indispensable. A decision table S is independent if all attributes in C are indispensable. *Let S = (U, A = C ∪ D) be a decision table.*

Rough Set Attribute Reduction (RSAR) provides a filter based tool by which knowledge may be extracted from a domain in a concise way; retaining the information content whilst reducing the amount of knowledge involved.

2.4 Reduct and Core

Let S = (U, A=C U D) be a decision table. A subset R of C is a reduct of C, if $POS_R(D) = POS_C(D)$ and S' = (U, RUD) is independent, ie., all attributes in R are indispensible in S'. Core of C is the set of attributes shared by all reducts of C. CORE(C) = ∩RED(C) where, RED(C) is the set of all reducts of C. The reduct is often used in the attribute selection process to eliminate redundant attributes towards decision making.

2.5 Correlation

Correlation define as a mutual relationship or connection between two or more things .The quantity *r*, called the *linear correlation coefficient*, measures the strength and the direction of a linear relationship between two variables. The linear correlation coefficient is sometimes referred to as the *Pearson product moment correlation coefficient* in honor of its developer Karl Pearson. The mathematical formula for its coefficient given by the formula

$$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right) - \left(\sum x\right)^2}\sqrt{n\left(\sum y^2\right) - \left(\sum y\right)^2}}$$

2.6 Goodness of fit

The **goodness of fit** of a statistical model describes how well it fits a set of observations. Measures of **goodness of fit** typically summarize the discrepancy between observed values and the values expected under the model in question.

2.7 Chi squared distribution

A **chi-squared test**, also referred to as **χ^2 test**, is any statistical hypothesis test in which the sampling distribution of the test statistic is a chi squared distribution when the null hypothesis is true. Also considered a chi-squared test is a test in which this is *asymptotically* true, meaning that the sampling distribution (if the null hypothesis is true) can be made to approximate a chi-squared distribution as closely as desired by making the sample size large enough. The chi-square (I) test is used to determine whether there is a significant difference between the expected frequencies and the observed frequencies in one or more categories. Do the number of individuals or objects that fall in each category differ significantly from the number you would expect? Is this difference between the expected and observed due to sampling variation, or is it a real difference

2.7.1 Further analysis of chi square test

Basic properties of chi squared goodness fit is that it is non symmetric in nature .How ever if the degrees of freedom increased it appears to be to be more symmetrical .It is right tailed one sided test. All expectation in chi squared test is greater than $1.E_I=np_i$ where n is the number samples considered p_i is the probability of i^{th} occurrence .Data selected at random there are two hypothesis null hypothesis and alternate hypothesis null hypothesis denoted by H_0 alternate hypothesis denoted by H_1. H_0 is the claim does follow the hypothesis and H_1 is the claim does not follow the hypothesis here H_1 is called the alternate hypothesis to H_0.If the test value found out to be K then K can be calculated by the formula $K=\sum(O_I-E_I)^2/ E_I$. Choice of significance level always satisfies type 1 error .

2.8 Different types of error-

1) Type 1 error-Rejecting a hypothesis even though it is true
2) Type 2 error-Accepting the hypothesis when it is false
3) Type 3 error-Rejecting a hypothesis correctly for wrong reason

3. BASIC IDEA

The basic idea for the proposed work is conceived from the general medical system. We initially consider 1000 samples, of jaundice cases and five conditional attributes such as such as Purities , Fatigue, Abdominal pain , Fever and Dark urine and two decision attribute positive and negative . The symptom regarding the dieses Jaundice collected from Doctor Pradeep Kumar Mishra MD

4. DATA REDUCTION

As the volume of data is increases with time it is difficult to know which symptoms are responsible for particular dieses. The basic objective of data reduction is to find the relevant

attributes that have all essential information about the data set. The process is illustrated by applying rough set concept on 20 samples which we get by using correlation technique applied on 1000 samples and 20 samples of jaundice case appears to be different . In this paper the five conditional attributes Purities ,Dark urine, Abdominal pain , Fever and Fatigue rename as a_1,a_2,a_3,a_4,a_5 the data's for these conditional attributes are low ,moderate and high rename as b_1,b_2,b_3 and decision attributes are positive , negative and rename as c_1,c_2 respectively .Application and analysis on the data set and rule generation being presented in the following tables . Table -1 is the initial table, and the process of analysis is present in the subsequent tables

Table-1

E	a_1	a_2	a_3	a_4	a_5	d
E_1	b_2	b_2	b_1	b_1	b_1	c_2
E_2	b_1	b_2	b_2	b_3	b_3	c_1
E_3	b_1	b_2	b_2	b_3	b_3	c_1
E_4	b_1	b_2	b_2	b_3	b_3	c_1
E_5	b_3	b_3	b_3	b_3	b_2	c_2
E_6	b_1	b_2	b_2	b_2	b_2	c_2
E_7	b_2	b_2	b_2	b_2	b_2	c_1
E_8	b_1	b_1	b_1	b_1	b_1	c_2
E_9	b_1	b_2	b_2	b_3	b_3	c_1
E_{10}	b_1	b_2	b_2	b_2	b_2	c_2
E_{11}	b_2	b_3	b_3	b_3	b_3	c_1
E_{12}	b_1	b_2	b_3	b_1	b_2	c_1
E_{13}	b_3	b_2	b_2	b_2	b_1	c_2
E_{14}	b_3	b_3	b_3	b_3	b_3	c_1
E_{15}	b_2	b_1	b_1	b_1	b_1	c_2
E_{16}	b_1	b_1	b_1	b_1	b_1	c_2
E_{17}	b_1	b_2	b_2	b_3	b_2	c_1
E_{18}	b_1	b_2	b_2	b_3	b_2	c_2
E_{19}	b_1	b_3	b_1	b_3	b_3	c_2
E_{20}	b_1	b_3	b_1	b_3	b_3	c_1

The decision table -1 , takes the initial values before finding the reduct looking at the data table it is found that entities $E_{17,}E_{18}$ ambiguous in nature and E_2,E_3, E_4 gives same result so both $E_{17,}E_{18}$ drop from the table and from E_2, E_3 , E_4 we keep the records of E_2 to avoid ambiguity. Table-2. derive from Table-1 shown below .

Reduced Table-2 from table-1

E	a_1	a_2	a_3	a_4	a_5	D
E_1	b_2	b_2	b_1	b_1	b_1	c_2
E_2	b_1	b_2	b_2	b_3	b_3	c_1
E_5	b_3	b_3	b_3	b_3	b_2	c_2
E_6	b_1	b_2	b_2	b_2	b_2	c_2
E_7	b_2	b_2	b_2	b_2	b_2	c_1
E_8	b_1	b_1	b_1	b_1	b_1	c_2
E_9	b_1	b_2	b_2	b_3	b_3	c_1
E_{10}	b_1	b_2	b_2	b_2	b_2	c_2
E_{11}	b_2	b_3	b_3	b_3	b_3	c_1
E_{12}	b_1	b_2	b_3	b_1	b_2	c_1
E_{13}	b_3	b_2	b_2	b_2	b_1	c_2
E_{14}	b_3	b_3	b_3	b_3	b_3	c_1
E_{15}	b_2	b_1	b_1	b_1	b_1	c_2
E_{16}	b_1	b_1	b_1	b_1	b_1	c_2
E_{19}	b_1	b_3	b_1	b_3	b_3	c_2
E_{20}	b_1	b_3	b_1	b_3	b_3	c_1

a. Lower Approximation

Lower Approximation is a description of the domain objects that areknown withdefinite belongto thesubsetof interest.

TheLowerApproximationSetofasetX,withregardtoRistheset ofallobjects,which can beclassifiedwith Xregarding R,that is denoted as R_L.

b. UpperApproximation

UpperApproximationisdefined to be possiblybelongtothe subsetof interest.TheUpperApproximationSetofasetXregardingRisth esetofall of objects which can be possibly classifiedwith Xregarding R . Denoted as R_U

c. Boundary Region(BR)

BoundaryRegionisdescriptionoftheobjectsthatofasetXregar dingRisthesetofallthe objects,whichcannot be classified neitherasXnor-XregardingR.IftheboundaryregionX=ϕ then the set is considered "Crisp", that is, exact in relation to R;

otherwise,iftheboundaryregionisasetX≠ϕthesetX"Rough"isc onsidered.In thattheboundary region is BR =R_U-R_L.

4.1 Analysis and classification result extraction

$E_{positive}$ ={E_4, E_5, E_7, E_9, E_{11}, E_{12}, E_{14} } ……. (1)

$E_{negative}$ ={E_1, E_2, E_6, E_8, E_{10}, E_{13}, E_{15},E_{16},E_{18} }…… (2)

$E(a_1)_{low}$={ E_4, E_6, E_8, E_9, E_{10}, E_{12}, E_{16},E_{17}, E_{18}}……(3)

$E(a_1)_{moderate}$={ E_1, E_2, E_7, E_{11}, E_{15} }……(4)

$E(a_1)_{high}$ ={ E_5, E_{13}, E_{14} }……..... (5)

The above result when compared with the positive cases $E(a_1)_{high}$ strength[11]
Found to be nil where as for negative cases of high $E(a_1)$ strength[11] 3/3 cent percent similarly for negative cases of moderate $E(a_1)$ strength[11] gives rise to be 1/5 about 20% , positive cases of low $E(a_1)$ strength[11] 6/9 about 66% basic observation gives rise is that a_1 attribute does not give any significant result similarly

$E(a_2)_{low}$={ E_8, E_{15}, E_{16} }……(6)

$E(a_2)_{moderate}$={ E_1, E_2, E_4,E_6,E_7,E_9,E_{10},E_{12},E_{13},E_{18} }……..(7)

$E(a_2)_{high}$ ={ E_5,E_{13},E_{14}}……(8)

Similar analysis strength[11] positive high a_2 will be 2/3 And for strength[11] for negative for low a_2 will be also 3/3=1 this gives a significant result so we keep the attribute a_2 for further classification . Now analyzing the a_3 we derive the following conclusion

$E(a_3)_{low}$={E_1,E_8,E_{12},E_{15},E_{16}}…….. (9)

$E(a_3)_{moderate}$={E_2,E_6,E_7,E_9,E_{10},E_{13}}……(10)

$E(a_3)_{high}$={E_5,E_{11},E_{12},E_{14} }……(11)

Finding the strength [11] high a_3 positive case will be 4/4=1 .And similarly for strength [11] for low a_3 cases will be gives negative result 4/5 .As in high a_3 we find large volume of positive jaundice case similarly in case of low a_3 we find large volume of negative case of jaundice. So the attribute a_3 has certain degree of importance, so keep a_3 for further classification

$E(a_4)_{low}$= {E_1,E_8,E_{12}, E_{15}, E_{16} }……(12)

$E(a_4)_{moderate}$= {E_6,E_7,E_{10},E_{13}}……(13)

$E(a_4)_{high}$= {E_2,E_5,E_9,E_{11},E_{14},E_{19} E_{20} }……(14)

Analyzing a_4 $E(a_4)_{low}$ negative strength[11] will be 4/5 similarly for $E(a_4)_{high}$ positive strength[11] cases will be about 4/7 so we cannot ignore this attribute now analyzing a_5 we have the following result as in low density of a_4 provide negative cases of jaundice

$E(a_5)_{low}$={E_1,E_8,E_{13},E_{15}, E_{16} }……(15)

$E(a_5)_{moderate}$= {E_5,E_6 ,E_7, E_{10}, E_{12} }……(16)

$E(a_5)_{high}$ = {E_2,E_9, E_{11}, E_{14}, E_{19},E_{20} }……..(17)

$E(a_5)_{high}$ strength[11] for positive case will be 5/6 similarly $E(a_5)_{low}$ strength[11] for negative cases found out to be 1 so we are having an ambiguity as because in positive case in high a_5 give a significant ratios in the contrast we are also a significant ratios in the negative strength[11] so we are ignoring this a_5 attribute from further classification , it is to be noted that the process of analyzing the data is just depends upon two cases that is the high and low cases , in case of not getting any significant result from two high and low case then we switch our analysis to moderate case in this context upon analyzing the data which collected we get all significant result from low and high case studies so we ignore the moderate case i.e from b_1 and b_3 .So we get the reduced table Table-3 from table-2 after dropping two attribute both a_1,a_5

Reduced Table-3 From Table-2

E	a_2	a_3	a_4	d
E_1	b_2	b_1	b_1	c_2
E_2	b_2	b_2	b_3	c_1
E_5	b_3	b_3	b_3	c_2
E_6	b_2	b_2	b_2	c_2
E_7	b_2	b_2	b_2	c_1
E_8	b_1	b_1	b_1	c_2
E_9	b_2	b_2	b_3	c_1
E_{10}	b_2	b_2	b_2	c_2
E_{11}	b_3	b_3	b_3	c_1
E_{12}	b_2	b_3	b_1	c_1
E_{13}	b_2	b_2	b_2	c_2
E_{14}	b_3	b_3	b_3	c_1
E_{15}	b_1	b_1	b_1	c_2
E_{16}	b_1	b_1	b_1	c_2
E_{19}	b_3	b_1	b_3	c_2
E_{20}	b_3	b_1	b_3	c_1

Upon analyzing table-3 we have the following result that is (E_6,E_7 ,E_{10}), (E_5,E_{11}), provide ambiguous result and (E_2,E_9), (E_8,E_{15},E_{16}) provide similar result so we drop (E_6,E_7 ,E_{10}), (E_5,E_{11}) form the table and keep single attribute E_2 and E_8 from the group (E_2,E_9), (E_8,E_{15},E_{16}) Table-4 given as follows

Reduced Table-4 From Table-3

E	a_2	a_3	a_4	d
E_1	b_2	b_1	b_1	c_2
E_2	b_2	b_2	b_3	c_1
E_5	b_3	b_3	b_3	c_2
E_8	b_1	b_1	b_1	c_2
E_{11}	b_3	b_3	b_3	c_1

E_{12}	b_2	b_3	b_1	c_1
E_{13}	b_2	b_2	b_2	c_2
E_{14}	b_3	b_3	b_3	c_1
E_{19}	b_3	b_1	b_3	c_2
E_{20}	b_3	b_1	b_3	c_1

Upon analyzing Table-4 we have the following conclusion i.e E_{14} and E_{20} provide certain degree of ambiguity i.e (E_{19},E_{20}) ambiguous in nature so in the subsequent table i.e Table-5 which derive from Table-4 we drop both entities E_{19} and E_{20}

Reduced Table-5 From Table-4

E	a_2	a_3	a_4	d
E_1	b_2	b_1	b_1	c_2
E_2	b_2	b_2	b_3	c_1
E_5	b_3	b_3	b_3	c_2
E_8	b_1	b_1	b_1	c_2
E_{11}	b_3	b_3	b_3	c_1
E_{12}	b_2	b_3	b_1	c_1
E_{13}	b_2	b_2	b_2	c_2
E_{14}	b_3	b_3	b_3	c_1

Again Further analysis provide that E_5,E_{11},E_{14} ambiguous in nature so we next drop E_5,E_{11},E_{14} from the table -5 and get the reduced table -6

Reduced Table-6 From Table-5

E	a_2	a_3	a_4	d
E_1	b_2	b_1	b_1	c_2
E_2	b_2	b_2	b_3	c_1
E_8	b_1	b_1	b_1	c_2
E_{12}	b_2	b_3	b_1	c_1
E_{13}	b_2	b_2	b_2	c_2

Further reduction of table-6 is not possible , so we derive the knowledge based algorithm looking at the reduced table-6 , the algorithm is present is entirely in knowledge based it completely based on the reduction table which we get after finding the reduct from the data set which we collected from medical resources

The algorithm is as follows

Rule

1 Dark urine Moderate ,low abdominal pain and low fever leads to negative hepatitis case
2 Dark urine Moderate ,moderate abdominal pain and high fever leads to positive hepatitis case
3 Dark urine low ,low abdominal pain and low fever leads to negative hepatitis case
4 Dark urine Moderate ,high abdominal pain and low fever leads to positive hepatitis case
5 Dark urine Moderate ,moderate abdominal pain and moderate fever leads to negative hepatitis case

From the above 5 step algorithm we get a conclusion that at least one of the entity is high to get a positive case of hepatitis so all three attributes are important in hepatitis detection i.e all three attribute Dark urine, abdominal pain and fever are responsible for hepatitis

Time complexly analysis- For finding the reduct we are comparing each record with another suppose there are n records then time complexity will be

$n+(n-1)+(n-2)+(n-3)+\ldots\ldots+1.=n(n-1)/2$ that is of $\Theta(n^2)$ that is the worst case analysis average case analysis of breaking the table $\Theta(nlgn)$ we are taking the average case analysis for dividing the table because on an average single table break down to half of it's size every time so the entire complexity will be $\Theta(nlgn+n^2)$

Statistical validation- For validate our findings we basically depends upon chi-square test for this purpose we consider we take a survey by taking data regarding the positive case and we are not focused on one medical centre to collect data we approached several hospital and the apply chi square test to validate our claim. . Chi square test- Expected15%,10%,15%,20%,30%,15% and the Observed samples are 25,14,34 45,62,20 so totaling these we have total of 200 samples so expected numbers of samples per each day as follows 30,20,30,40,60,30 . We then apply chi square distribution to verify our result assuming that H_0 is our hypothesis that is correct H_1 as alternate hypothesis that is not correct , Then we expect sample in six cases as

chi squared estimation formula is $\sum(O_i-E_i)^2/ E_i$ where i=0,1,2,3,4,5 so the calculated as follows

$X^2=(25-30)^2/20+(14-20)^2/20+(34-30)^2/30+(45-40)^2/40+(62-60)^2/60+(20-30)^2/30$

$X^2=25/20+36/20+16/30+25/40+4/60+100/30=7.60$ the tabular values we have with degree of freedom 5 we get result 11.04

This is far below the tabular value, so we accept the hypothesis that our claim is correct .

Future work- This theory can be extended to the field of entertainment , Small scale business sector

5. REFERENCES

[1] Cios, K., W. Pedrycz and R. Swiniarski (1998). *Data Mining Methods for Knowledge Discovery*. Kluwer Academic

[2] Wolf, S., H. Oliver, S. Herbert and M. Michael (2000). Intelligent data mining for medical quality management

[3] Se-Ho, Ch., and P. Rockett (2002). The training of neural classifiers with condensed datasets. *SMCB*, **32**(2), 202–206.,

[4] Setiono, R. (2000). Generating concise and accurate classification rules for breast cancer diagnosis. *Artificial Intelligence in Medicine*, **18**(3), 205–219

[5] Cheeseman, P., and J. Stutz (1996). Bayesian classification (AutoClass): theory and results. In U.M.Fayyad

[6] Grzymala–Busse, J., Z. Pawlak, R. Slowinski and W. Ziarko (1999). Rough sets. *Communications of the ACM*

[7] Hassanien, A.E. (2003). Classification and feature selection of breast cancer data based on decision tree algorithm

[8] Parido, A., and P. Bonelli (1993). A new approach to fuzzy classifier systems. In *Proceedings of the FifthInternational Conference on Genetic Algorithms*. pp. 223–230

[9] Lin, T.Y., and N. Cercone (1997). *Rough Sets and Data Mining*. Kluwer Academic Publishers.Ning, S., H. Xiaohua, W. Ziarko and N. Cercone (1994). A generalized rough sets model. In *Proceedings ofthe 3rd Pacific Rim International Conference on Artificial Intelligence*, Vol. 431. Beijing, China. Int. Acad.Publishers. pp. 437–443.

[10] Pawlak, Z. (1991). *Rough Sets-Theoretical Aspect of Reasoning about Data*. Kluwer Academic Publishers. Pawlak, Z., J. Grzymala–Busse, R. Slowinski, W. Ziarko (1995). Rough sets. *Communications of the ACM.*

[11] Renu Vashist Prof M.L Garg Rule Generation based on Reduct and Core :A rough set approach International Journal of Computer Application(0975-887) Vol 29 September -2011 Page 1-4